Technologies for Optical Countermeasures XVII; and High-Power Lasers: Technology and Systems, Platforms, Effects IV
Proceedings Volume 11539; SPIE Security + Defence, 21-25 September 2020, Online UK
DOI:10.1117/12.2574099

Countermeasure Leveraging Optical Attractor Kits (CLOAK): interpretational disruption of a visual-based workflow

Marco Zaccaria Di Fraia*, Lounis Chermak

Centre for Electronic Warfare Information and Cyber, Cranfield University, The Defence Academy Of The UK, Shrivenham, SN6 8LA, UK

ABSTRACT

Due to their negligible cost, small energy footprint, compact size and passive nature, cameras are emerging as one of the most appealing sensing approaches for the realization of fully autonomous intelligent mobile platforms. In defence contexts, passive sensors, such as cameras, represent an important asset due to the absence of a detectable external operational signature – with at most some radiation generated by their components. This characteristic, however, makes targeting them a quite daunting task, as their active neutralization requires pinning a small angular diameter moving at a high speed.

In this paper we introduce an interpretational countermeasure acting against autonomous platforms relying on feature-based optical workflows. We classify our approach as an interpretational disruption because it exploits the heuristics of the model used by the on-board artificial intelligence to interpret the available data. To remove the struggle of accurately pinpointing such an imperceptible target, our approach consists in passively corrupting, from a perception point of view, the whole environment with a small, sparse set of physical observables.

The concrete design of these systems is developed from the response of a feature detector of interest. We define an optical attractor as the collection of pixels inducing an exceptionally strong response for a target feature detector. We also define a physical object inducing these pixel structures for defense purposes as a CLOAK: Countermeasure Leveraging Optical Attractor Kits. Using optical attractors, any optical based algorithm relying on features extraction can potentially be disrupted, in a completely passive and nondestructive fashion.

Keywords: Artificial Intelligence, Optical countermeasures, UAVs, Features detection, Extreme values, Computer Vision

1. INTRODUCTION

In recent years the diffusion of commercial drones has highlighted important vulnerabilities in the current conception of defensive systems. Even ignoring military-specific developments like the reconnaissance-specific nanoUAV PD-100 Black Hornet[1], inexpensive, widely available UAVs (Unmanned Aerial Vehicles) are enabling weak actors to deploy with minimal effort unconventional offensive or intelligence payloads beyond defensive perimeters. Remarkable examples of this are the use of drones as IEDs carriers during the 2018 Caracas drone attack[2] or in Iraq and Afghanistan[3].

In particular, our interest lies in off-the-shelf, small-scale, highly pervasive platforms capable of carrying a malicious payload without the need for large scale deployment and control structures. Specifically, with reference to NATO standard ATP-3.3.8.1[4] these would be Class I drones, generally Mini, or, at most, Small.

Let us, within the scope of this paper, focus on target locations which do not represent soft targets. Through this assumption we can expect the actor not to be able to operate the platform in Visual Line of Sight (VLOS). Small dimensions, moreover, make it impossible to employ dedicated beyond line of sight technologies. Additionally, assuming that the operator has not the support of a dedicated GNSS, which is plausible because of the weak actor hypothesis, and that the area is geofenced to commercial services and jammed, the UAV has to possess some autonomous capability to be able to complete its mission.

*Marco.Di-Fraia@cranfield.ac.uk

By assuming an on-board GN&C system capable of autonomously handling GNSS-denied operations we are considering a platform capable of bridging the external operational context to the estimated ego-motion. While the latter can be measured by sensors not concerned with the structure of the surrounding scenario (e.g. accelerometers, magnetometers, etc.) the former requires some form of awareness of the external environment, generally introduced through electro-optical sensors. By taking into account all the constraints on the platform, camera sensors are considered as the most likely choice for this purpose. Indeed cameras offer an affordable, power efficient, and light-weight, passive-sensing solution, capable of generating an extensive amount of texture and structure based information of the surroundings scene. Exploitation of these data enables cameras to turn into a remarkable navigation instrument, capable of providing localisation information through various algorithms. Consequently, we are considering here a commercial UAV equipped with camera-enabled autonomy.

Now, given all the above characteristics it is evident that countering these threats means confronting a virtually ubiquitous and potentially present in a large number (swarm) menace. Therefore, instead of an active seek-and-destroy (or seek-and-blind) approach we suggest restructuring the environment in such a way that, passively, the offensive platform finds it hard or impossible to operate within it.

In particular, we present a way to disruptively manipulate the processes of an on-board artificial intelligence employing visual based processes by exploiting its operational structure. Namely, we observe that most computer vision processes begin with the detection of significant, highly distinguishable regions of the image, known as features. Thus, the whole subsequent workflow depends on interpreting and manipulating these computational objects. Hence, we aim to exploit the interpretational process of the artificial intelligence, by diverting the focus of the detection towards specific regions and then operating on this restricted set of observables.

This is achieved by performing an augmentation and design process on the environment. The scope of this process is to induce regions in the computational representation of the scene which strongly excite the feature detector. This allows to have the certainty that the features which the algorithm assigns a higher score to are entities over which the defender has full control. Hence they represent important access point to the complete downstream process; thus manipulating the real objects inducing them equals propagating in the interpretation process.

In ultimate analysis, this is a work exploiting the extrema of the detection process. We call the pixel patches inducing these extrema for specific detectors Optical Attractors (OAs), and the physical objects associated to them Optical Attractor Kits (OAK). Finally, the complete countermeasure approach operating on this principle is termed, as per title, Countermeasures Leveraging Optical Attractor Kits (CLOAKs).

While future developments will be focusing on what we call the “inverse autonomy problem”, which is estimating the architecture of the on-board AI from its apparent behaviour so as to then deploy the appropriate countermeasure, the architecture is assumed here to be known, through intelligence reports or as descending from standardized controllers.

2. METHODS AND BACKGROUND

As stated above, we are concerned with hindering the feature-based computer vision processes of aerial autonomous and intelligent light-weight intruders. These algorithmic pipelines are generally mission-enabling, as would be the case for optical navigation[5]. We believe that in order to counter this class of ubiquitous threats, it is paramount to develop passive optical countermeasures. Namely, the deployed defense should be largely context-intrinsic, i.e. part of the operational scenario, and not requiring triggers or activation. From an historic perspective this can be seen as a translation of disruptive patterns, mainly employed during the first World War on naval vessels (dazzle camouflage)[6], targeting, this time, AI operators, rather than human ones.

Let us for now focus on a single intruding platform, employing only one feature detector algorithm. In a previous work[7] we observed that under certain circumstances there is an undesired by-effect caused by the introduction of a restricted number of environmental augmentation elements. Namely, being these observables specifically designed to induce a strong response in detectors they have the potential to completely drive the observation process of these algorithms, sometimes with disruptive effects. This translates to the concept that a small set of physical objects can be propagated within the computational detection and interpretation process to manipulate, set back or even neutralise the operations of an autonomous platform. In particular, this can happen in at least two ways.

For a given image I_k acquired during ordinary operative conditions, a feature detector generally associates a score p_{jk} to each detected feature $X_{jk,}$ which characterizes the quality of the detection.

After the M features satisfying the boundary conditions have been retrieved, let us order the scores in a decreasing order. Let the number $i = 1, \dots, M; i \in \mathbb{N}$ indicate the rank of the feature with the score q_{ik}. Let us call the feature with the highest score "pivot", and let its score be indicated with $Q_{0k} = q_{1k}$. Experimentally is it possible to observe that it is generally possible to find three parameters such that $q_{ik}(i)$ is well approximated by a power law over a range of values

$$q_{ik}(i) \quad = c_1 i^{c_2} + c_3 \tag{1}$$

A noteworthy exception to this is represented by images containing fractals structures where the distribution tends to be stair-like. As an example, running a SURF (Speeded Up Robust Feature)[8] detection with all default values over the built-in MATLAB demo image termed "concordorthophoto.png" it is possible to retrieve 22435 points, that when ordered have the distribution shown in Figure 1. Figure 2 shows the 80 strongest SURF features detected within the underlying image.

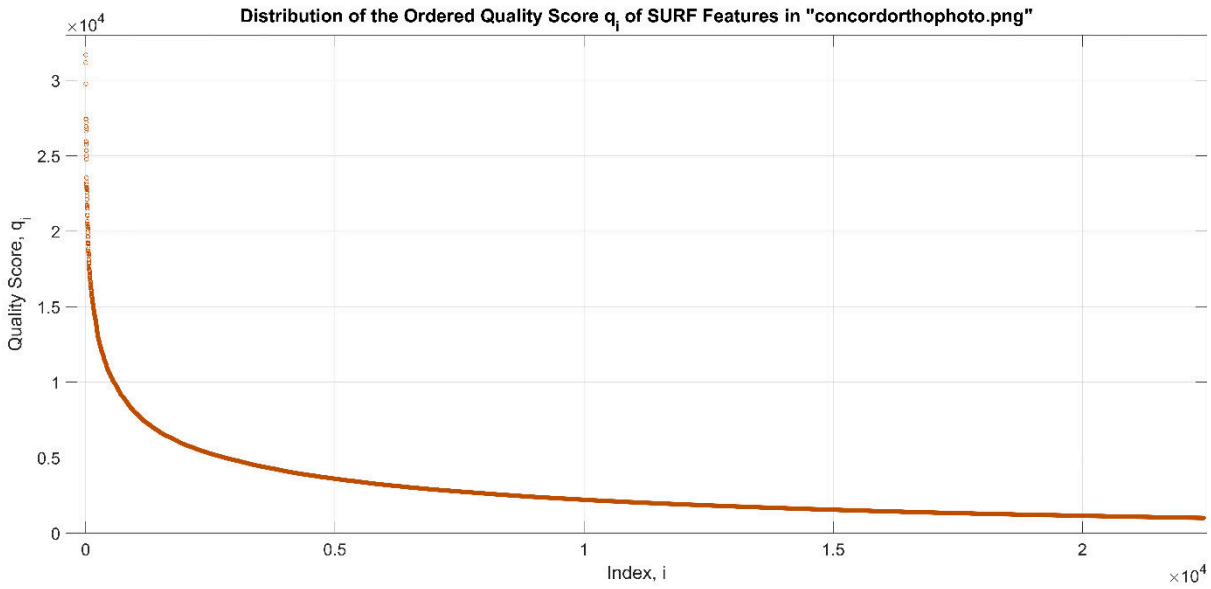

Figure 1. Distribution of the Ordered Quality Score q_i of SURF Features in "concordorthophoto.png"

Scores and pivots are important because their values drive the acceptance or rejection of pixel patches as features. Let us illustrate this through the MATLAB 2019b implementation of two local feature detectors, SURF and Harris-Stephens corners[9]. It is worth noting that conceptually the two detector look for different objects (blobs and corners, respectively), both from an algebraic and data perspective; however this will be not a concern within this work.

Let us drop the subscript k, and let us define a threshold T. Without adding computational safeguards, within MATLAB, Harris- Stephens' threshold operates in such a way that a feature is accepted as such if $Q_0 > p_j > T_H Q_0$, with $T_H \ll 1$. On the other hand, for surf $T_s \gg 1$ and a generic feature j is accepted if $Q_0 > p_j > T_S$. We call a feature with a rejection process of the first type Pivot-Led (PL) and features with a rejection process of the second way Pivot-Independent (PI).

Figure 2. The 80 strongest SURF features (green crosses and circles) detected within the underlying image.

At this stage, it is finally possible to introduce the first two approaches enabling the disruption of a visual based process. First, a PL process gives the ability to shift and manipulate the whole rejection interval by introducing an OA. It can be added that in the case where the rest of the environment is built in such a way that the number and quality of the remaining features is kept low, it is effectively possible to consistently induce a scarcity of detected features. We call this approach "starving". On the other hand, since the lower boundary in a PI process cannot be externally manipulated, it is not possible to exploit it in the same way. This is why it requires a more complex approach, which we defined "herding".

Whereas starving is a static technique, herding, is a dynamic technique, requiring the scene to change between images. That being said, it may still be considered to be a passive system when there is no activation prompted by an external signal. Neglecting for now their thermal effect as only solutions in the visible band are considered, as a first iteration we can consider CLOAKs to be screens capable of showing a grayscale image. This enables us to consider different OAs interchangeably, and is consistent with a dynamic application. Therefore, ultimately, herding consists in continuously moving OAs between the screens, structuring their motion pattern in such a way that the same descriptor ends up applying to objects which are continuously shifting position in space. As these features represent by design the ones with the highest scores, the detector rarely rejects them, thus constantly integrating disruption-inducing elements in the process.

3. RESULTS

Regardless of the rejection approach, there are two main processes in the deployment of a CLOAK. These are the "Extraction" and "Injection" phases. The former, extraction, refers to determining optical attractors for the given

configuration of a feature detector, through either heuristics, or optimization techniques. Afterwards, in the injection stage, these are introduced within the field of view of the sensor, in such a way that process disruption is maximized.

Within the scope of this study a naïve full version of this workflow is developed and analysed, again in MATLAB (v. R2019b). In particular we are interested in presenting two results: that it is possible to retrieve optical attractors, and that they have a significant effect when introduced in an image. We will focus our attention on SURF with default parameters, and limit our scope to static tests.

For the extraction stage optimization processes known as genetic algorithms[10] were used to "evolve" an ideal optimal response to the analysed parametrization of the feature detector.

Figures 3a, 3b and 3c show the outputs of this process with the retrieved optical attractors for square pixel patches of different sizes (75x75, 105x105, and 255x255). OAs can be clearly seen in all the three figures as the cross-like structure surrounded by a continuous circle of the opposite colour.

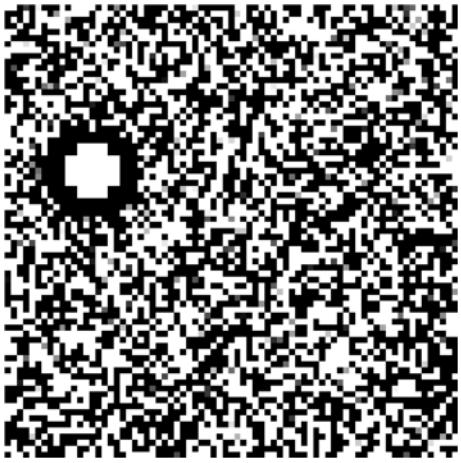

Figure 3a. The Optical Attractor emerged in the 75x75 square.

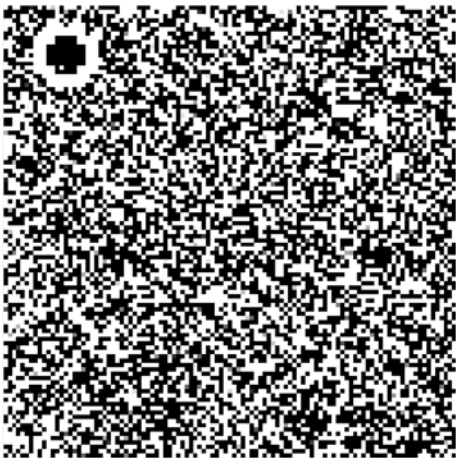

Figure 3b. The Optical Attractor emerged in the 105x105 square.

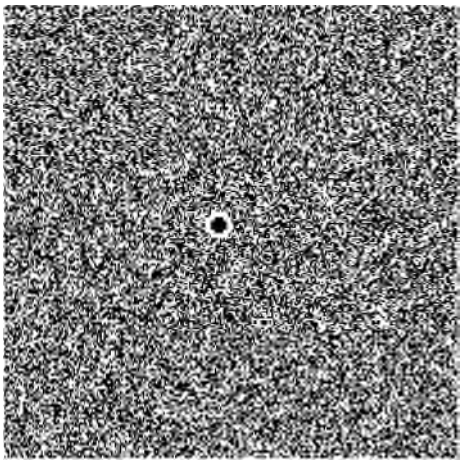

Figure 3c. The Optical Attractor emerged in the 255x255 square.

It is worth noting that although it might appear that the OAs shrink with a larger image side in pixels, this impression is simply an artefact of this representation, where pictures with an increasing side length in pixels were kept at a fixed metric size.

Fig. 4 on the other hand, shows the q_i obtained from the three images above using a SURF detector with values left to default. It is immediately possible to see the three OAs values the in the upper left corner.

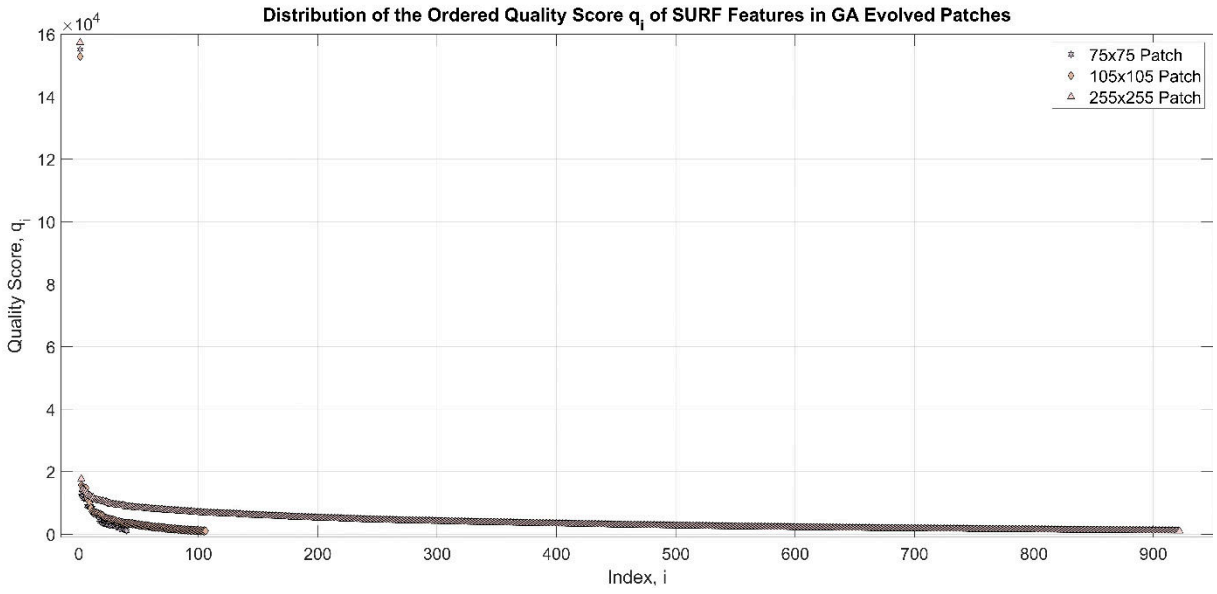

Figure 4. Distribution of the Ordered Quality Score q_i of SURF Features in GA Evolved Patches.

Let us now extract the 15x15 pixels patch containing the OA from Fig. 3c, which as per Fig. 4 is the one with the highest Q_0 between the three, and obtain a transferable OA. This is an approximation, because the shape of the OA appears to be circular, while the selected patch is a square. However it can be seen that this choice does not appear to affect the Q_0 in a significant way.

By performing a first proof of concept test, pasting in random places within the built-in test images mentioned also for Fig. 1, "concordorthophoto.png" the square OA it is possible to observe two things:

a. The Q_0 value remains essentially constant in whatever picture and position the OA is found, unless…

b. The OA is placed on the edge of the image. In that case Q_0 reverts to one of the picture not containing the OA. Hence, from a statistical point of view, this second phenomenon is correlated to the image size.

From Fig. 5 it is possible to see that out of 1000 Monte Carlo runs, 98.4% of the times the behaviour falls under the first category.

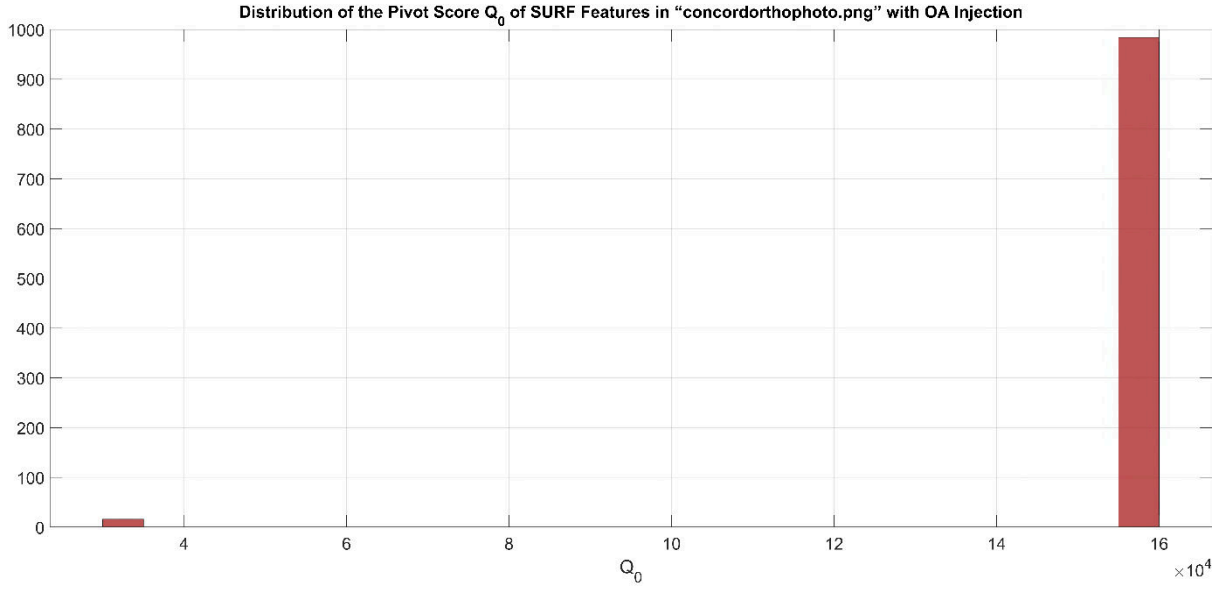

Figure 5. Distribution of the Pivot Score Q_0 of SURF Features in "concordorthophoto.png" with OA Injection

4. CONCLUSIONS AND FUTURE WORK

In this paper we presented Countermeasures Leveraging Optical Attractor Kits (CLOAKs), a set of EO Countermeasures targeting and exploiting visual based interpretation processes of on-board AIs for autonomous platform, in particular UAVs. In addition to that it was necessary to introduce ancillary concepts which the development of these countermeasures is based on, in particular those connected to extrema values in computer vision. Lastly, we recognized two important phases in the operative deployment of these objects, extraction and injection, and presented tools and results enabling the development and analysis of CLOAKs.

The results appear promising, and deserving further investigation.

There are multiple lines of work that are expected to descend from this study. The first, obvious, one concerns retrieving optimization-based optical attractors for all the feature detectors for which this problem is well-defined. Secondly, structuring an accurate mathematical search of the highest possible extrema, to validate the genetic search, or, eventually, find improved solutions, and developing a score to compare the environment to the optimal feature shape. Thirdly, implementing dynamic tests in order to experimentally validate the proposed herding behavior. Lastly, examining the solvability of the inverse autonomy problem is an important objective: given observations on the motion, the observable payload and of the target, determining the structure of the on-board AI.

ACKNOWLEDGMENTS

Marco's PhD is sponsored by Thales Alenia Space. Marco would like to thank Giancarmine Fasano for being such an exceptional and inspiring teacher in the field of UAVs. This paper would have not existed without his teachings.

REFERENCES

[1] Nano UAS PD-100 Black Hornet, https://bssholland.com/products/nano-uas-pd-100-black-hornet/
[2] Koettl, C., and B. Marcolini. "A closer look at the drone attack on maduro in venezuela." The New York Times (2018).
[3] Rossiter, Ash. "Drone usage by militant groups: exploring variation in adoption." Defense & Security Analysis 34.2 (2018): 113-126.
[4] NATO, "ATP-3.3.8.1," 2016
[5] Nistér, David, Oleg Naroditsky, and James Bergen. "Visual odometry." Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.. Vol. 1. Ieee, 2004.
[6] Titterton, David H. "A review of the development of optical countermeasures." Technologies for Optical Countermeasures. Vol. 5615. International Society for Optics and Photonics, 2004.
[7] Di Fraia, Marco Zaccaria, et al. "NAV-Landmarks: Deployable 3D Infrastructures to Enable CubeSats Navigation Near Asteroids." 2020 IEEE Aerospace Conference. IEEE, 2020.
[8] detectSURFFeatures (function), https://uk.mathworks.com/help/vision/ref/detectsurffeatures.html
[9] detectHarrisFeatures (function), https://uk.mathworks.com/help/vision/ref/detectharrisfeatures.html
[10] Genetic Algorithm, https://uk.mathworks.com/discovery/genetic-algorithm.html